

УДК 004.891

Б. Хомів, С. Лупенко

(Тернопільський національний технічний університет імені Івана Пулюя)

ЗАСТОСУВАННЯ ВАГОВИХ КОЕФІЦІЄНТІВ ПРИ ОЦІНЮВАННІ ОПІНІЇ ТЕКСТОВОЇ ІНФОРМАЦІЇ У WEB-ДОКУМЕНТАХ

В доповіді розглядається актуальна науково-технічна задача застосування вагових коефіцієнтів при оцінюванні опінії текстової інформації. Наведено основні процедури порівняння компонент об'єкта та об'єктів в цілому, котрі використовуються на даний час, проаналізовано їхні відмінності та обґрунтовано важливість застосування вагових коефіцієнтів при оцінюванні опінії текстової інформації у web-документах. При аналізі текстової інформації формується граф ієрархії об'єкт, компонент, підкомпонент, атрибут з відповідними значеннями (позитивний, негативний, нейтральний, можуть бути й інші варіації, залежно від обраного методу класифікації). Наприклад, проаналізувавши речення «Хороша фотокамера, ціна та якість відмінні, задоволений ергономікою» об'єктом виступить слово «фотокамера», компонентами об'єкта слова «ціна, якість, ергономіка», атрибутами компонент слова «хороший, відмінні, задоволений».

Згідно попереднього огляду виявлено, що такі системи, як Opinion Observer, OPINE, Review Seer, Red Opal, Web Fountain, Opinion Miner, Bing Shopping, Google Products, Quark Shop при резюмуванні результатів не зводять компоненти об'єкта до єдиного інтегрального показника або ж зводять їх не враховуючи вагових коефіцієнтів.

Тобто, при інтегральному оцінюванні опінії для об'єкту «фотокамера» його компоненти такі як «ціна» та «якість» матимуть однакову вагу, що є не цілком коректним з точки зору покупця. Проте, якщо ввести вагові коефіцієнти між такими компонентами як «батарея», «зображення», «ергономіка» та «ціна», то може виявитися, що компонента «ціна» важливіша за компоненту «батарея», а компонента «ергономіка» важливіша за компоненту «ціна» при порівнянні двох об'єктів (рис 1).

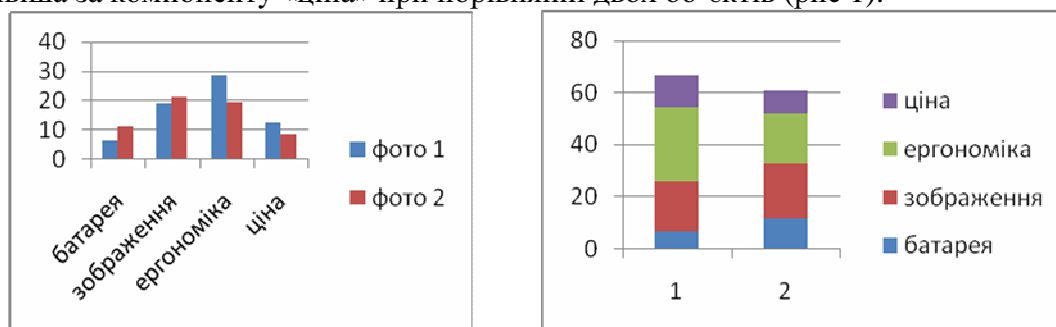


Рис 1. Діаграми зважених компонент

Проаналізувавши методи підрахунку позитивних та негативних висловлювань, було виявлено наступні техніки порівняння компонент об'єктів та об'єктів між собою:

1. Ранжування по кількості згадувань із застосуванням лінійної регресії.
2. Ранжування по кількості згадувань із застосуванням нечіткої логіки.
3. Ранжування по кількості покупок із застосуванням економетричних моделей.

Також у доповіді звернено увагу на задачі застосування вагових коефіцієнтів при порівнянні не лише компонент об'єкта, але й окремих висловлювань, наприклад «хороший дисплей», «непоганий дисплей», методи поділу коментарів на певну кількість класів.

Застосування вагових коефіцієнтів уможливить отримати більш релевантний, більш точний та якісніший результат при оцінюванні опінії текстової інформації.